Equal-Probability Partitioning from Random Projections

Random projection systems frequently produce approximately Gaussian-distributed coordinates due to central-limit effects. This enables several simple methods for constructing balanced discrete partitions and symbolic encodings from continuous vectors.

These constructions are especially relevant for:

- associative memory,
- locality-sensitive hashing,
- randomized indexing,
- distributed representations,
- vector databases,
- and neural feature quantization.

---

1. Binary Partition from a Single Gaussian Coordinate

Assume a projection coordinate:

$$x \sim \mathcal{N}(0,1)$$

The simplest balanced partition is:

$$b = \mathbf{1}_{x \geq 0}$$

which produces:

- b=0 with probability 0.5,
- b=1 with probability 0.5.

This is effectively a random hyperplane partition.

Multiple independent coordinates yield multiple bits:

$b_1, b_2, ..., b_n$

producing:

2^n

equiprobable regions under independence assumptions.

This construction is closely related to:

- SimHash,
- random hyperplane LSH,
- binary embeddings,
- and sign-based neural representations.

---

2. Equal-Probability Quantization Using the Gaussian CDF

A more general approach uses the Gaussian cumulative distribution function:

$u=\Phi(x)$

where:

where:

$$\Phi(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$$

For:

$$x \sim \mathcal{N}(0,1)$$

the transformed variable satisfies:

$$u \sim \text{Uniform}(0,1)$$

Discrete equal-probability bins are then obtained by:

$$k = \lfloor Mu \rfloor$$

giving:

$$k \in \{0, \ldots, M-1\}$$

with each symbol occurring with equal probability 1/M.

Example:

- M=256 gives balanced 8-bit symbolic states.

Unlike linear quantization, this avoids:

- overcrowded central bins,
- sparse tail bins,
- and poor entropy utilization.

The resulting symbols maximize occupancy balance and entropy efficiency.

---

3. Angular Partitioning from Two Coordinates

Using two approximately independent Gaussian coordinates:

$$x, y \sim \mathcal{N}(0,1)$$

the angle:

$$\theta = \operatorname{atan2}(y, x)$$

is uniformly distributed over:

$$[-\pi, \pi)$$

This provides another natural equal-probability coordinate.

Uniform angular bins are obtained by:

$$k = \left\lfloor M \frac{\theta + \pi}{2\pi} \right\rfloor$$

producing:

- rotationally balanced partitions,
- phase-like symbolic encodings,
- and circular quantization schemes.

This method differs from independent sign partitioning because it preserves:

- relative directional structure,
- rotational geometry,
- and pairwise phase relationships.

---

4. Relation to Random Projection Systems

Orthogonal or randomized mixing transforms such as the Walsh–Hadamard transform naturally produce approximately Gaussian coordinates when applied recursively or to sufficiently distributed inputs.

This gives a generic pipeline:

$x \to Hx \to$ Gaussian-like coordinates $\to$ equal-probability partitioning

where H may be:

- a Hadamard transform,
- random orthogonal transform,
- or randomized projection matrix.

The resulting symbolic states are:

- balanced,
- entropy efficient,
- and globally mixed.

---

5. Geometric Interpretation

The constructions correspond to different partition geometries:

Method	**Geometry**	**Output**
sign((x))	hyperplane split	binary bit
multiple signs	orthant partition	binary code
Gaussian CDF	equal-mass intervals	balanced integer
atan2((y,x))	angular sectors	phase symbol

These provide different tradeoffs between:

- locality preservation,
- entropy balance,
- rotational invariance,
- and implementation simplicity.

---

6. Potential Applications

Possible applications include:

- associative memory addressing,
- vector hashing,
- neural symbolic compression,
- balanced quantization,
- randomized routing,
- distributed indexing,
- probabilistic state encoding,
- and recursive dynamical systems.

The equal-partition viewpoint emphasizes probability balance rather than geometric uniformity, which can improve symbol utilization and collision behavior in high-dimensional randomized systems.